

修 士 論 文 の 和 文 要 旨

| | | | |
|--|------------------------------|------|---------|
| 研究科・専攻 | 大学院 情報理工学研究科 情報学専攻 博士前期課程 | | |
| 氏 名 | 宮崎 夏美 | 学籍番号 | 1730103 |
| 論 文 題 目 | ソーシャルネットワークのプロファイリングと個人特定の研究 | | |
| <p>要 旨</p> <p>近年、SNS におけるプライバシーの侵害が問題となっている。プライバシーリスクを明確化するために、先行研究では同一人物による複数のアカウントを特定する手法が提案されている。本研究ではリスクをより明確化するために実世界の個人を特定する手法に焦点を当て、機械学習を用いて Twitter アカウントと就職希望者の履歴書を照合することで、匿名の発言の投稿者を見つける先行研究を取り上げた。これらの信頼性を高めるとともに、SNS から投稿者が持つ属性をプロファイリングしその精度を確認、分析することで推定されやすい人物の持つ性質を考察した。</p> <p>プロファイリングについては、Twitter の投稿文から投稿者の属性を推定した。教師データについてはデータセットを正例、負例共に最大 100 人と先行研究より増加させ、また被験者を 78 人に変更し、属性数を大幅に増やして精度を確認した。結果、最も識別精度の高い属性カテゴリは性別であり、学歴、帰省先住所と続いた。ここで複数の被験者について、履歴書に記入されていなかった属性の推定値が高かったことから、投稿者の潜在的な属性の推定についての可能性を示した。</p> <p>投稿者の隠れた属性の推定については、申告されていないが機械学習によって本人のものと推定された属性が実際に正しいのかを分析し、それにより潜在的な属性を見つけることができる可能性を示した。機械学習による属性推定とアンケートによる事後申告により、14 人中 10 人については事前に申告がない属性についても、本人のものと推定された属性のアンケートスコアが実際に高くなっていた。個人特定については、78 人中 29 人を正しく特定し、78 人中 65 人を 1 割(すなわち上位 7 位)以内に絞り込むことができた。これは先行研究の結果(30 人中 6 人を特定)を上回る。また学内被験者では 27 人中 12 人、学外被験者では 51 人中 27 人を正しく特定した。</p> | | | |

電気通信大学情報理工学研究科

平成 30 年度修士論文

ソーシャルネットワークの
プロファイリングと個人特定の研究

学籍番号 1730103

氏名 宮崎 夏美

情報学専攻 セキュリティ情報学コース

主任指導教員 吉浦 裕 教授

指導教員 内海 彰 教授

提出日 平成 31 年 1 月 29 日(月)

概要

近年、SNS におけるプライバシーの侵害が問題となっている。プライバシーリスクを明確化するために、先行研究では同一人物による複数のアカウントを特定する手法が提案されている。本研究ではリスクをより明確化するために実世界の個人を特定する手法に焦点を当て、機械学習を用いて **Twitter** アカウントと就職希望者の履歴書を照合することで、匿名の発言の投稿者を見つける先行研究を取り上げた。この先行研究の信頼性を高めるとともに、SNS から投稿者が持つ属性をプロファイリングし、その精度を確認、分析することで推定されやすい人物の持つ性質を考察した。

プロファイリングについては、**Twitter** の投稿文から投稿者の属性を推定した。履歴書に記入された属性の中から、「性別」、「年代」、「帰省先住所」、「学歴」、「職種」、「得意科目」、「クラブ活動・サークル・趣味」、「資格」の 8 カテゴリに対して属性推定を行った。教師データについてはデータセットを正例、負例共に最大 100 人と先行研究より増加させ、また被験者を 78 人に変更し、属性数を大幅に増やして精度を確認した。結果、最も識別精度の高い属性カテゴリは性別であり、学歴、帰省先住所と続いた。反対に最も識別精度の低い属性カテゴリは年代であり、資格、得意科目と続いた。ここで複数の被験者について、履歴書に記入されていなかった属性の推定値が高かったことから、投稿者の潜在的な属性の推定についての可能性を示した。

投稿者の隠れた属性の推定については、申告されていないが機械学習によって本人のものと推定された属性が実際に正しいのかを分析し、それにより潜在的な属性を見つけることができる可能性を示した。機械学習による属性推定とアンケートによる事後申告により、14 人中 10 人については事前に申告がない属性についても、本人のものと推定された属性のアンケートスコアが実際に高くなっていた。

個人特定については、78 人中 29 人を正しく特定し、78 人中 65 人を 1 割(すなわち上位 7 位)以内に絞り込むことができた。これは先行研究の結果(30 人中 6 人を正しく特定、30 人中 12 人を 1 割以内に特定)を上回る。また学内被験者では 27 人中 12 人を正しく特定し、27 人中 16 人を 1 割(すなわち上位 2 位)以内に、学外被験者では 51 人中 27 人を正しく特定し、51 人中 40 人を 1 割(すなわち上位 5 位)以内に絞り込むことができた。

内容

| | |
|------------------------------|----|
| 第 1 章 序論..... | 1 |
| 第 2 章 先行研究..... | 2 |
| 2.1 SNS からのプロファイリング..... | 2 |
| 2.2 匿名アカウントから電子世界の個人を特定..... | 2 |
| 2.3 匿名アカウントから実世界の個人を特定..... | 3 |
| 第 3 章 データセット..... | 5 |
| 3.1 履歴書情報..... | 5 |
| 3.2 Twitter アカウント..... | 5 |
| 第 4 章 プロファイリング..... | 7 |
| 4.1 属性識別器の作成..... | 7 |
| 4.2 プロファイリング結果と考察..... | 7 |
| 第 5 章 個人特定..... | 12 |
| 5.1 履歴書識別器の生成..... | 12 |
| 5.2 個人特定の結果と考察..... | 14 |
| 5.3 推定されやすい人物..... | 15 |
| 第 6 章 隠れた属性の推定..... | 17 |
| 6.1 目的..... | 17 |
| 6.2 実験方法..... | 17 |
| 6.3 推定の結果と考察..... | 18 |
| 第 7 章 結論..... | 21 |
| 7.1 まとめ..... | 21 |
| 7.2 今後の課題..... | 21 |
| 謝辞..... | 22 |
| 参考文献..... | 22 |

| | |
|----------|----|
| 付録 | 24 |
|----------|----|

第 1 章 序論

昨今、ソーシャルネットワークサービス(以下 SNS とする)がコミュニケーションの手段として広く普及し、人と人、人と企業などの間で意思疎通を円滑かつ、活発にしている。また SNS 上でのコミュニケーションは、日常的な意思疎通に限らず、匿名での内部告発などにも利用されることがある。

一方、SNS 上の投稿は多くの場合に不特定多数が閲覧できる状態になっている。そのため、投稿者の意思にかかわらずプライバシー情報が漏洩したり、漏洩した情報から個人が特定されたりする可能性がある。Facebook や Twitter といった SNS サービスでは、このようなプライバシー情報の漏洩を防ぐために、発言内容の公開範囲を設定できる機能が用意されている。しかしこの機能を使用することはユーザの負担であり、情報発信のたびに適切な設定を行う必要がある。加えて SNS の本来の目的である「コミュニケーションの楽しみ」を損なう可能性があるなど、利用者が不便に感じたり楽しみが軽減してしまう部分が多い。そこで多くのユーザは、名前、年齢、住所、所属、顔などの個人識別情報を、投稿文やプロフィール上で省略・変更・隠蔽をして、友人のみがその人物を認識できるように匿名化を実施し、情報の漏洩を避けている。

しかし、これらの対策にも関わらず、同一人物に関する複数の情報が入手されることで、その人物のプライバシーが侵害される危険性が懸念されている。特に、SNS で個人の特定を回避した情報が別のメディアの情報と比較、照合されることでその匿名性を失うとの報告が具体的な手法とともに明らかにされている[1]。

プライバシーへの攻撃には 2 種類あり、一つはアカウントの投稿者が分かっており、投稿者の属性や行動を推定するプロファイリング、もう一つは匿名のアカウントの投稿からその投稿者を推定する個人特定である。しかし、従来のプロファイリングは飲酒や病気など限られた種類の推定しか行っておらず、個人特定も同一人物の複数のアカウントや投稿を見つけるものであり、これは実世界の個人を特定しているのではない。実世界の個人を特定する先行研究もある[2]が、被験者が少なく偏っている。

プロファイリングと個人特定は密接な繋がりがあり、匿名の SNS アカウントのプロファイリングによって投稿者の属性を推定することで、その匿名の個人が誰であるかの特定が可能となる。本研究では、匿名の SNS アカウントをプロファイリングし、その結果に基づいて、アカウントの所有者を特定する。プロファイリングは履歴書に相当する数百項目について実施する。また先行研究の問題であった被験者の偏りを減らし、人数を増やすことで信頼性を高める。

これによって偏りのない多くの被験者の協力を得て、SNS の匿名アカウントからのプロファイリングおよび個人特定の精度を明らかにするとともに、プロファイリングしやすい属性、特定しやすい個人の持つ性質を明らかにする。

第 2 章 先行研究

2.1 SNS からのプロファイリング

SNS 上から，個人情報を取り出すための様々な攻撃方法が報告されている．初期の研究では主に，キーワードの照合のような単純な手法で利用者の個人情報を推定することに焦点が当てられてきた．その後，2011 年に Mao らは，ナイーブベイズ分類器や SVM(Support Vector Machine)といった学習アルゴリズムを利用することで，飲酒状態で投稿された投稿を 84%の精度で，また旅行や病気に関する投稿を 76%の精度で検出できると報告している[3]．この研究に使用された訓練データは，個人情報を含む投稿と含まない投稿であり，そのラベル付けは人手で行われていた．2012 年に Kótyuk と Buttyan は，本人や友人のプロフィールおよび投稿文で開示された情報から，年齢，性別，未婚・既婚などの関係をニューラルネットワークによって学習することで，開示された情報から開示されていない情報を推定できるとしている[4]．2014 年に Caliskan-Islam らは，ナイーブベイズ分類器とアダプテストを用いて，ユーザの個人情報の漏洩度合を 3 段階に分類した．

2.2 匿名アカウントから電子世界の個人を特定

近年では，対象の SNS アカウントやその投稿を，別の SNS と照合するといった手法が提案されている．2009 年に Narayanan らは，サブグラフマッチングにより，Twitter と Flickr を利用する同一ユーザを 12%の誤り率で特定する手法を提案している[1]．そして 2012 年には，SVM や線形判別機などの機械学習アルゴリズムを使用することで，同一ユーザが投稿した複数のブログを推定することができるとしている[5]．2010 年に Polakis らは，Facebook の情報を利用し，SNS 上のユーザ名と実在する電子メールアドレスを照合することができたと報告している[6]．具体的には，Twitter と Facebook の情報を活用することで，43%のユーザプロフィールと実際のメールアドレスとの対応付けを行うことができた．2012 年に Goga らは，位置情報，タイムスタンプおよび書式情報を解析することで，異なる SNS(Yelp, Twitter, Flickr)を利用する同一ユーザを特定する手法を提案している[7]．同年，Narayanan は SVM や線形判別分析を含むいくつかの学習アルゴリズムを用いて，同一人物によるブログの投稿を特定した[8]．2014 年に Almishari らは，ナイーブベイズ分類器を使用し，同一のユーザが使用している Twitter アカウントを特定した[9]．これらの研究は同じアカウント同士の投稿文の照合であり，あくまでも電子の世界での個人を特定しているのみである．

2.3 匿名アカウントから実世界の個人を特定

橋本らによって、SNS のアカウントと履歴書を機械学習によって照合し、匿名アカウントの人物を特定するシステムが提案されている[2]。この手法では、ある企業が自社に対する誹謗・中傷などを発言したアカウントを特定する状況を想定している。具体的には企業の持つ履歴書を利用し、発言した匿名アカウントの人物を複数の履歴書の中から特定する(図 2.1)。注目するアカウントと複数の履歴書との間の類似度をそれぞれ計算し、類似度の最も高い履歴書をアカウントのユーザであると判断する。

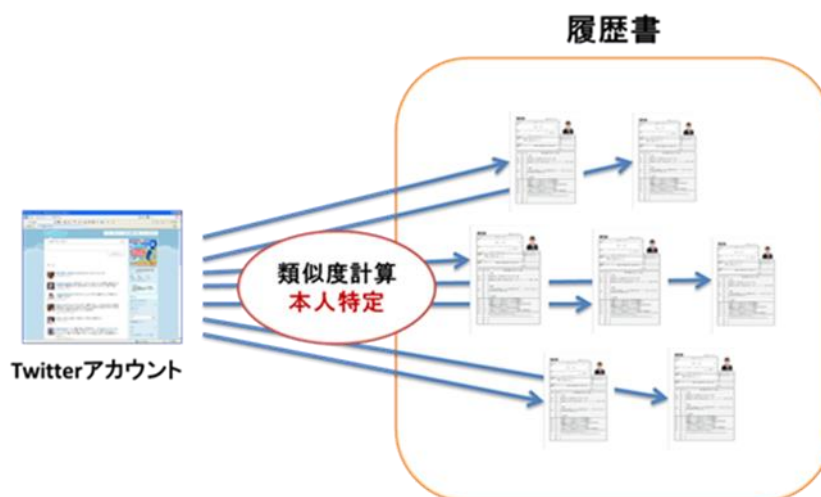


図 2.1 履歴書との照合による発言者特定システム概念

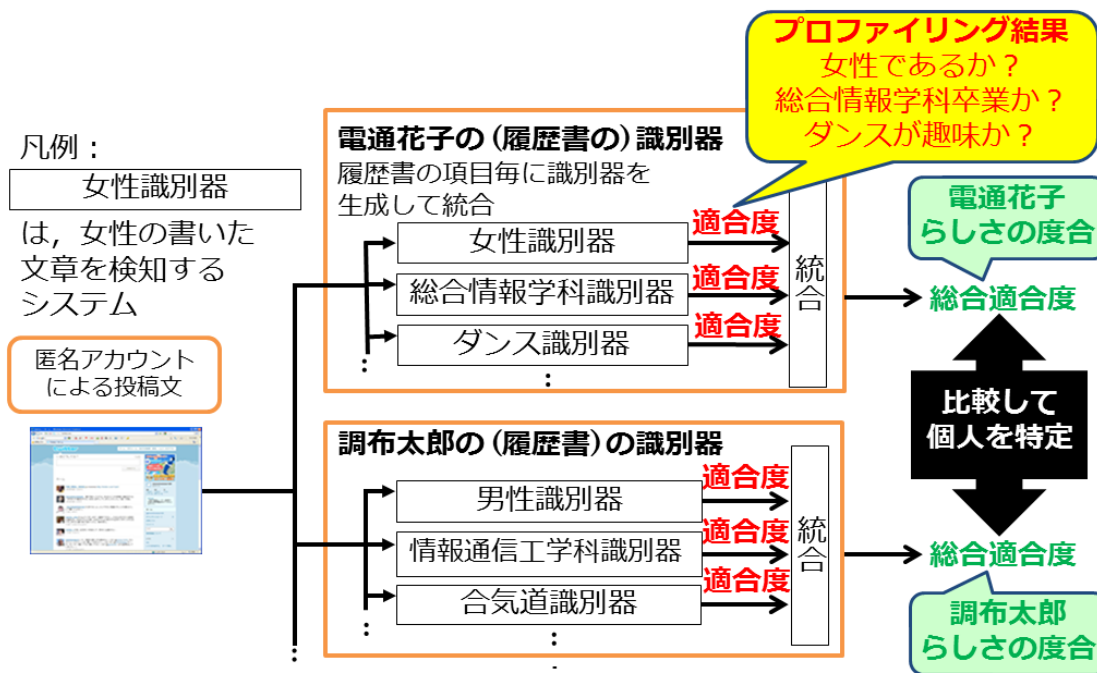


図 2.2 機械学習による識別の基本概念

まず対象となる **Twitter** アカウントの人物を特定するために、複数の履歴書をもとに属性識別器を作成する。図 2.2 に示すように、識別器はすべての履歴書に記載されている多数の属性値(例えば、性別=女、現住所=東京、趣味=テニス、etc.)それぞれに対して作成し、その属性値を有するか否かで判別を行うものとしている。

次に履歴書毎の識別器を作成する。例えば、電通花子という人物の履歴書に「性別:女性」, 「学科:総合情報」など、という記載があった場合、これらの識別器を組み合わせることによって電通花子の識別器を作成する。

なおこの研究で用いた属性は「住所」「学科」「趣味」「帰省先」「得意科目」「資格」「性別」の 7 種類、計 119 個である。

属性識別器の作成方法は、各属性値を持つアカウントを **Twitter** のプロフィール検索サービスなどによって複数入手し、これを正例とする。またその属性値を持たない複数のアカウントも同様に入手し、これを負例とする。これらを教師データとし、**SVM** や **Random Forest** などといった機械学習アルゴリズムを用いて属性識別器を学習する。

作成したそれぞれの識別器に対し、対象の **Twitter** アカウントのつぶやきをテストデータとして入力し、それぞれの識別器から出力されたスコアを統合する。統合の方式として平均または乗算を用いて、アカウントの人物が履歴書の人物そのものであるかを判定する。そのスコアが最も高い履歴書の人物を、対象の匿名アカウントの人物として判定する(図 2.2)。

この研究では、30 人分の **Twitter** アカウントのうち、6 人が本人の履歴書と正しく特定された。しかし、この研究には問題点があると考えられる。

まず、この実験の被験者である電気通信大学生徒 30 人というのは、被験者としては数が少なく、また被験者の属性にも偏りがある。そのため評価結果の信頼性が低いと考えられる。

また、この研究は 30 人分の **Twitter** アカウントのうち、本人の履歴書と正しく特定できたのは 6 人という低い結果である。この結果に対してどのような考察ができるかということについては言及されておらず、プライバシーのリスクを十分に示していない。

第3章 データセット

3.1 履歴書情報

被験者実験により，電気通信大学に在籍する学生 27 名に加えて，電気通信大学以外に在籍する学生や社会人 51 名，計 78 名の履歴書と Twitter アカウントを収集した．収集した履歴書情報は以下のとおりである．

- (1) 性別
- (2) 年代(18-22, 23-26, 27-30・・・4 歳刻み)
- (3) 現住所(市・区・郡まで)
- (4) 帰省先住所(市・区・郡まで)
- (5) 学歴
- (6) 職種
- (7) 得意科目
- (8) 長所・特徴（自己 PR）
- (9) クラブ活動・サークル・趣味
- (10) 資格

正例となるアカウントが収集できない等の理由より上記全ての情報は利用しておらず，実際に評価に用いたデータは(1) 性別，(2)年代，(4) 帰省先住所，(5) 学歴，(6) 職種，(7) 得意科目，(9) クラブ活動・サークル・趣味，(10) 資格 の計 8 項目である．ただし，これから 8 項目のうち一部が欠落している履歴書もある．

得意科目や趣味など，人によっては複数の属性が該当する項目もあるため，属性の数は履歴書によって異なる．一人当たりの属性の数は最低で 6 項目，最大で 28 項目であり，平均は 11.14 項目となっている．標準偏差は 3.7 である．

3.2 Twitter アカウント

本実験では，Twitter への投稿数が被験者は 500 ツイート，教師データは 1000 ツイート以上のアカウントを対象とし，被験者・教師データのどちらのアカウントも，アカウントあたり最新の 3000 ツイートまでを用いた．被験者 78 人のツイート数は最低で 535 ツイート，最大で 3222 ツイートであり，平均は 2351.85 ツイートとなっている．標準偏差は 809.70 である．

本実験において，正例を集めるためにはそれぞれの属性・属性値を持つ Twitter アカウントを検索する必要がある．その方法としてツイプロ[10]を用い，教師データとなるアカウントを入手した．ここで，教師データに使用したアカウントは 1 つの属性・属性値のペアにつき最大で 100 アカウントである．ただし，一般的に珍しい属性値(例:小型船舶操縦免許など)

は 100 アカウントも収集できない場合があり，その場合は最低 10 アカウントあればその属性値を用いることとしている．

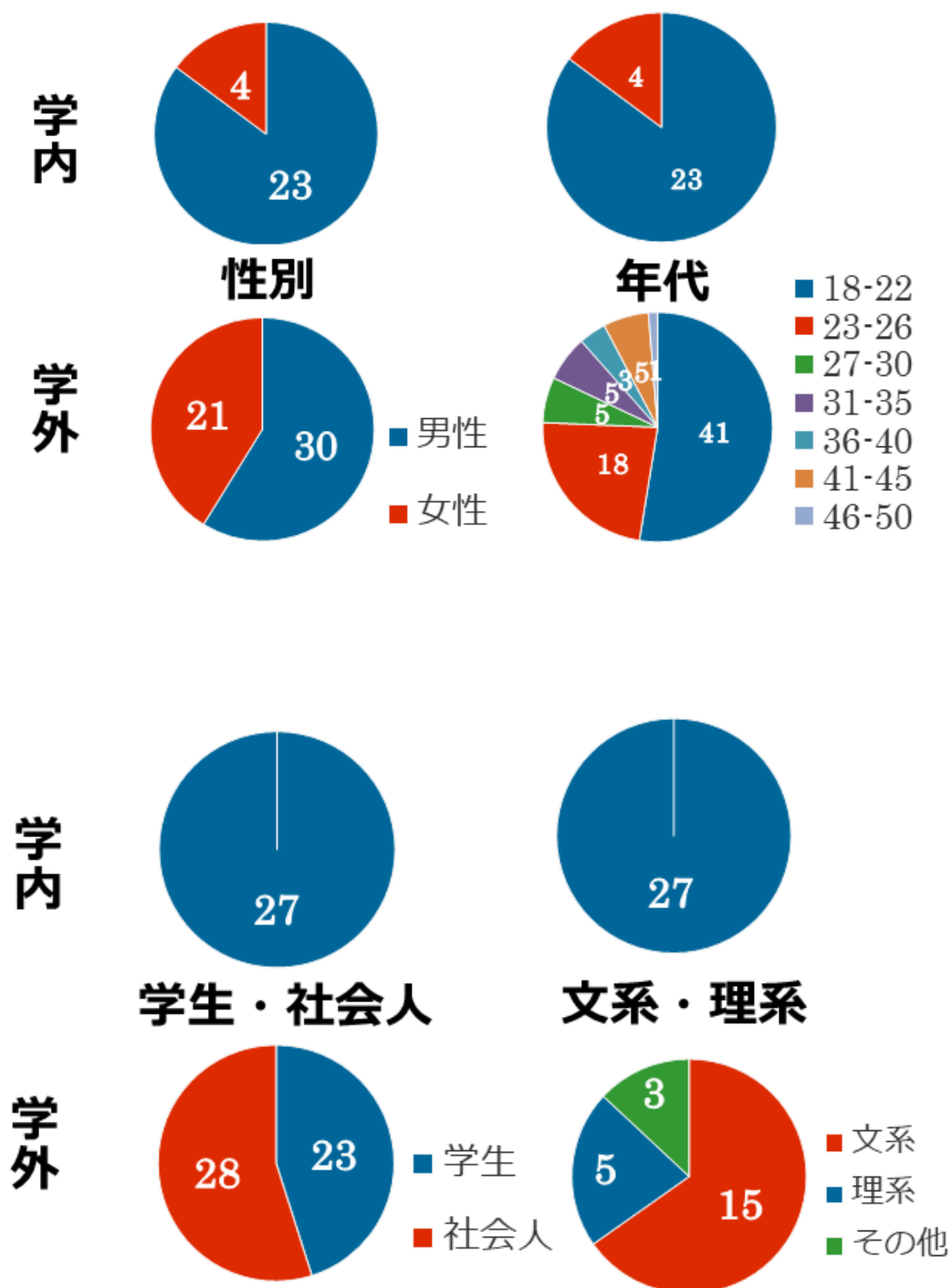


図 3.1 被験者の内訳

第4章 プロファイリング

4.1 属性識別器の作成

実験では、属性毎にツイプロで取得したユーザのつぶやきを用いた。具体的な方法は、識別器を作成したい属性をプロフィールにもつアカウントを正例、もたないアカウントを負例とし、ツイプロを用いて収集する。それら1アカウント当たり最新のものから過去3000件までのつぶやきを収集し、形態素解析機 MeCab によって解析する。

形態素解析とは、テキストを形態素まで分割し、各形態素における品詞を判別する作業であり、形態素と言語学の用語で、それ以上分解したら意味を成さなくなるまで分割された、語の最小の単位である。この形態素の出現頻度を特徴ベクトルとし、それを入力データとして機械学習を行った。

先行研究ではこの学習データに用いるアカウントの数を正例、負例それぞれ最大30アカウントで行っていた。しかし一般に学習データ数が多いほうが識別の精度が向上するため、人数を増やし最大で100アカウントの正例、負例を用いたところ、精度が向上した。そのため本実験では正例、負例それぞれ最大100アカウントを学習データとする。

形態素解析に使用する MeCab 用の辞書として、IPA 辞書と NEologd というものがある。NEologd は IPA 辞書に多数の Web 上の言語資源から得た新語を追加することでカスタマイズされたもので、定期的に更新を行うことで新語に対応している。この NEologd と IPA 辞書のどちらを用いた識別器の精度が良いかを検証したところ、IPA 辞書の方が優れていたため、本実験では IPA 辞書を用いている。

なお学習アルゴリズムは DMLC[11]によって開発されている Gradient Tree Boosting を実行するライブラリ XGBoost[12]を用い、特徴量としては Bag of words (単語の出現頻度)を使用した。

4.2 プロファイリング結果と考察

今回プロファイリングに用いた属性の例を示す(表 4.1)。カテゴリとは履歴書においてどの情報にあたるかのカテゴリライズ、項目の例とは今回の被験者が属している属性の一例であり、項目数とはそのカテゴリにおける属性が被験者全体でいくつあったのかの数である。識別精度の高い属性とは、実際にその属性を持つ被験者のスコアが高くなった属性の一例を示している。

一部の被験者のスコアの一例を表 4.2 に示す。下線が引いてあるものは実際にその属性を持つ被験者のスコアを示している。

最も識別精度の高いカテゴリは性別であり、学歴、帰省先住所と続いた。最も精度の低いカテゴリは年代であり、資格、得意科目と続いた。識別精度の基準については後述する。

性別は個人情報の中でも推定されやすい属性であるので妥当な結果である。学歴や帰省先住所の識別精度が高いというのは、SNS 上においてその属性が被験者個人にとって(意識

的か無意識的に関わらず), あまりリスクの高いプライバシー情報と考えられていない, かつ自己を表ことができるユニークな情報として発信しやすいものであると考えられているのではないかと考察できる. 例えば, 帰省先住所の中でも東京などあまり珍しくない(人口の多い)属性は識別精度が低く, また学歴についても, A 大学と B 大学といったような大学の識別精度は高かったが, A 大学 α 学科と A 大学 β 学科といったような学科の識別精度は低かった.

年代の識別精度が低かった理由は, 本実験に用いた年齢の区分が小さい, もしくは文章に年代が表れにくいことが考えられる. 資格や得意科目については, これらについてユーザがツイッター上でつぶやく必要を感じていないことが考えられる. 資格に関してはあまりにユニークなものだとプライバシーのリスクを感じ意図的に発信していない可能性も存在する.

表 4.1 属性の例

| カテゴリ | 項目の例 | 項目数 | 識別精度の高い属性 |
|---------------|---|-----|--|
| 性別 | 男・女 | 2 | 男・女 |
| 年代 | 18-22・23-26・27-30・・・ | 7 | |
| 帰省先住所 | 愛知県・新潟県・熊本県・・・ | 57 | 愛知県・新潟県・熊本県・・・ |
| 学歴 | 電気通信大学・・・ | 20 | |
| 通勤先・職種 | | 26 | |
| 得意科目 | Web セキュリティ・アルゴリズム・通信システム・電磁気学・・・ | 30 | Web セキュリティ・・・ |
| クラブ活動・サークル・趣味 | アーチェリー・アウトドア・アクションゲーム・アマチュア無線・クラシックギター・ゲームセンター・ハロプロ・ポケモン・ラグビー部・ラクロス部・横浜 DeNA ベイスターズ・軽音楽部・書道・将棋・人狼ゲーム・吹奏楽部サークル・声優・登山・放送研究会・留学・・・ | 174 | アーチェリー・アウトドア・アマチュア無線・ゲームセンター・ハロプロ・ポケモン・ラグビー部・ラクロス部・横浜 DeNA ベイスターズ・将棋・登山・放送研究会・・・ |
| 資格 | 海上無線・歴史検定・・・ | 24 | |

表 4.2 属性識別器スコアの一部

| | 性別 | | 年齢 | | 帰省先住所 | |
|--------|-----------------|-----------------|-----------------|-----------------|------------------|------------------|
| | 男 | 女 | 18-22 | 23-26 | 埼玉県 A 市 | 埼玉県 B 市 |
| 被験者 1 | <u>0.999491</u> | 0.000508 | <u>0.07145</u> | 0.452265 | <u>0.9963329</u> | 0.9499788 |
| 被験者 2 | <u>0.99668</u> | 0.00332 | <u>0.568197</u> | 0.583949 | 0.0401079 | 0.0035348 |
| 被験者 3 | <u>0.976975</u> | 0.023025 | <u>0.022415</u> | 0.711347 | 0.0473873 | 0.0134592 |
| 被験者 4 | <u>0.985342</u> | 0.014658 | <u>0.359974</u> | 0.05046 | 0.0144113 | 0.553514 |
| 被験者 5 | <u>0.978814</u> | 0.021186 | <u>0.017261</u> | 0.057622 | 0.5193191 | 0.0040804 |
| 被験者 6 | <u>0.999063</u> | 0.000937 | 0.005669 | <u>0.065433</u> | 0.0149693 | 0.0105171 |
| 被験者 7 | <u>0.996149</u> | 0.003851 | <u>0.034488</u> | 0.939049 | 0.0313453 | 0.0034176 |
| 被験者 8 | <u>0.92969</u> | 0.07031 | <u>0.339534</u> | 0.968809 | 0.036622 | 0.0035604 |
| 被験者 9 | <u>0.996976</u> | 0.003024 | <u>0.269929</u> | 0.125967 | 0.1433513 | 0.009207 |
| 被験者 10 | <u>0.994946</u> | 0.005054 | 0.194858 | <u>0.816119</u> | 0.0150451 | 0.006207 |
| 被験者 11 | 0.572584 | <u>0.427416</u> | <u>0.938805</u> | 0.773064 | 0.0333756 | <u>0.9505212</u> |
| 被験者 12 | <u>0.972512</u> | 0.027488 | <u>0.643045</u> | 0.517789 | 0.0098698 | 0.028918 |

| | 学歴 | | 資格 |
|--------|-----------------|-----------------|----------|
| | 電気通信大学 II 類 | 電気通信大学総合情報学科 | 英検 |
| 被験者 1 | 0.964861 | <u>0.964861</u> | 0.982245 |
| 被験者 2 | <u>0.964861</u> | 0.964861 | 0.990228 |
| 被験者 3 | <u>0.035139</u> | 0.964861 | 0.984859 |
| 被験者 4 | 0.035139 | <u>0.035139</u> | 0.850149 |
| 被験者 5 | <u>0.964861</u> | 0.964861 | 0.990107 |
| 被験者 6 | 0.964861 | <u>0.964861</u> | 0.990106 |
| 被験者 7 | <u>0.964861</u> | 0.964861 | 0.912451 |
| 被験者 8 | 0.964861 | <u>0.964861</u> | 0.810888 |
| 被験者 9 | <u>0.964861</u> | 0.964861 | 0.773549 |
| 被験者 10 | 0.964861 | <u>0.964861</u> | 0.962318 |
| 被験者 11 | <u>0.964861</u> | 0.964861 | 0.181146 |
| 被験者 12 | 0.035139 | <u>0.964861</u> | 0.837895 |

| | 得意科目 | | 趣味 | | |
|--------|--------------------|-----------------|----------------|-----------------|-------------------|
| | 電磁気学 | 統計学 | ギター | アウトドア | 将棋 |
| 被験者 1 | 0.946441472 | 0.842011 | 0.003796 | <u>0.992896</u> | 0.17590185 |
| 被験者 2 | 0.826048076 | 0.891354 | 0.005011 | 0.33884 | 0.00693983 |
| 被験者 3 | 0.303416908 | 0.751036 | 0.04345 | 0.683938 | 0.00539208 |
| 被験者 4 | 0.179372653 | 0.689696 | 0.034995 | 0.897145 | 0.00918772 |
| 被験者 5 | 0.968943834 | <u>0.848978</u> | 0.00317 | 0.053821 | <u>0.13826333</u> |
| 被験者 6 | 0.963331163 | 0.898875 | 0.006539 | 0.893033 | 0.02333358 |
| 被験者 7 | 0.588758707 | 0.738795 | <u>0.99733</u> | 0.443365 | 0.82408625 |
| 被験者 8 | 0.461890519 | 0.160672 | 0.036705 | 0.648482 | 0.00578586 |
| 被験者 9 | <u>0.951130629</u> | 0.829191 | 0.00706 | 0.936702 | <u>0.68964982</u> |
| 被験者 10 | 0.94460988 | 0.235798 | 0.011699 | 0.19905 | 0.01005645 |
| 被験者 11 | 0.79578948 | 0.343645 | 0.002327 | 0.096907 | 0.0084399 |
| 被験者 12 | 0.910803437 | 0.74463 | 0.292225 | 0.790898 | 0.00321814 |

属性値識別器の精度は以下の方法を用いて数値化した。

α という属性について、78 人の被験者のうち、該当者が N 人、非該当者が M ($M=78-N$) 人とする。

- (1) 78 アカウントを α の属性識別器で処理し、スコア順に 78 アカウントを整列する。
- (2) 該当者の各アカウントの重みを $78/N$ 、非該当者の各アカウントの重みを $78/M$ とする。
- (3) 閾値を L とする。整列した 78 アカウントのうち $1\sim L$ 番を、属性 α を持つ被験者、 $(L+1)\sim 78$ 番を、属性 α を持たない被験者と判定する。
- (4) (3)において判定が異なるアカウントの重みの和を誤り値とする。
誤り値 = ($1\sim L$ 番の中の非該当アカウントの重みの和) + ($(L+1)\sim 78$ 番の中の該当アカウントの重みの和)
- (5) L を 1 から 78 まで変更し、各々誤り値を求め、最小の誤り値を、属性値識別器の誤り率とする。この誤り率が小さいほど識別精度が高い。

表 4.3 属性識別器誤り率の一例

| 属性値 | | 誤り率 |
|-------|----------|----------|
| 性別 | 男 | 12.3034 |
| | 女 | 12.3034 |
| 年代 | 18-22 | 65.29993 |
| | 23-26 | 55.03333 |
| 帰省先住所 | 愛媛県 D 市 | 0 |
| | 神奈川県 E 市 | 7.090909 |
| | 東京都 F 区 | 70.81579 |
| 学歴 | G 大学 | 7.090909 |
| | H 大学 | 0 |
| 資格 | 漢検 | 33.58333 |
| | 教員免許 | 60.55263 |
| 得意科目 | 数学 | 68.4 |
| | 統計学 | 5.131579 |
| | 体育 | 66.08333 |
| 趣味 | アーチェリー | 0 |
| | ギター | 14.56 |
| | ファッション | 49.63636 |
| | フットサル | 0 |

第 5 章 個人特定

5.1 履歴書識別器の生成

本研究の目的は，SNS の投稿内容と履歴書の情報を照合し，直接的な個人特定を行うことである．そのため，4.1 節で作成した属性識別器を組み合わせることで履歴書識別器を作成する．

具体的な例を示す(図 5.1)．とある人物 Alice の履歴書が以下の様であったとする．この時，Alice は「性別：女」「現住所：東京」「帰省先：北海道」「趣味：ダンス」「趣味：ピアノ」という属性値を所有しているものとし，それぞれの識別器を作成する．つまり，Alice の履歴書では 5 つの識別器を作成する．1 つの項目内に複数の属性値がある場合（例えば，「趣味」の項目に「ダンス」と「ピアノ」の 2 つが記入されている），それぞれの属性値毎に識別器を作成する．

これら 5 つの識別器に特定したい Twitter アカウントの投稿を入力し，得られたスコアを平均することで統合し Alice の履歴書に対する照合度合を算出する．

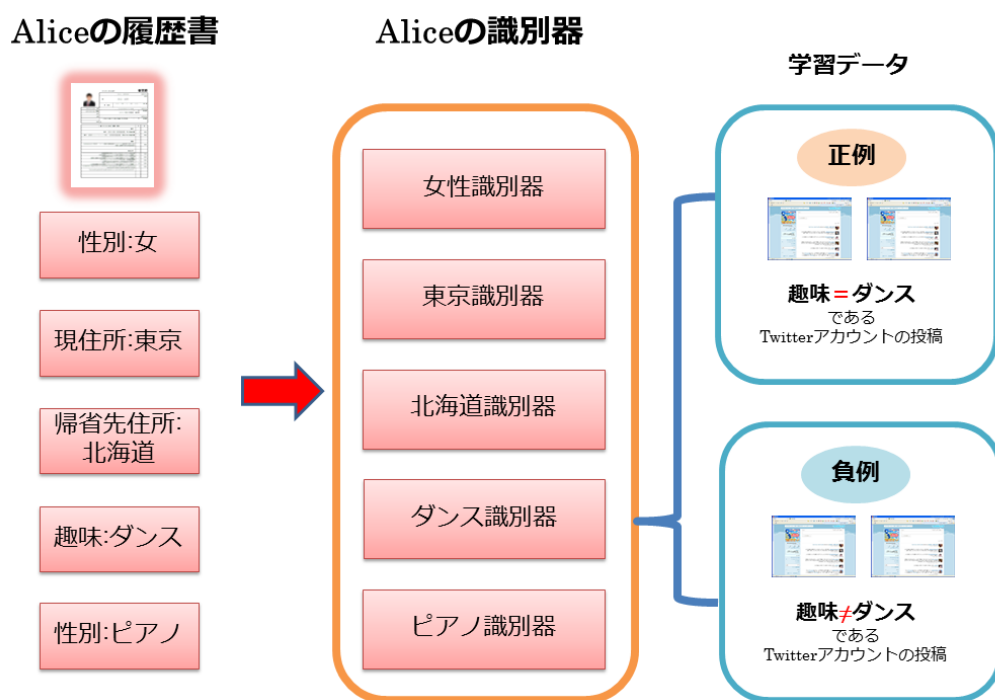


図 5.1 履歴書識別器の作成方法

本研究では被験者の履歴書をもとに 78 人分の履歴書識別器を作成し、最も照合度が高い履歴書を入力した Twitter アカウント本人の履歴書とする (図 5.2)。

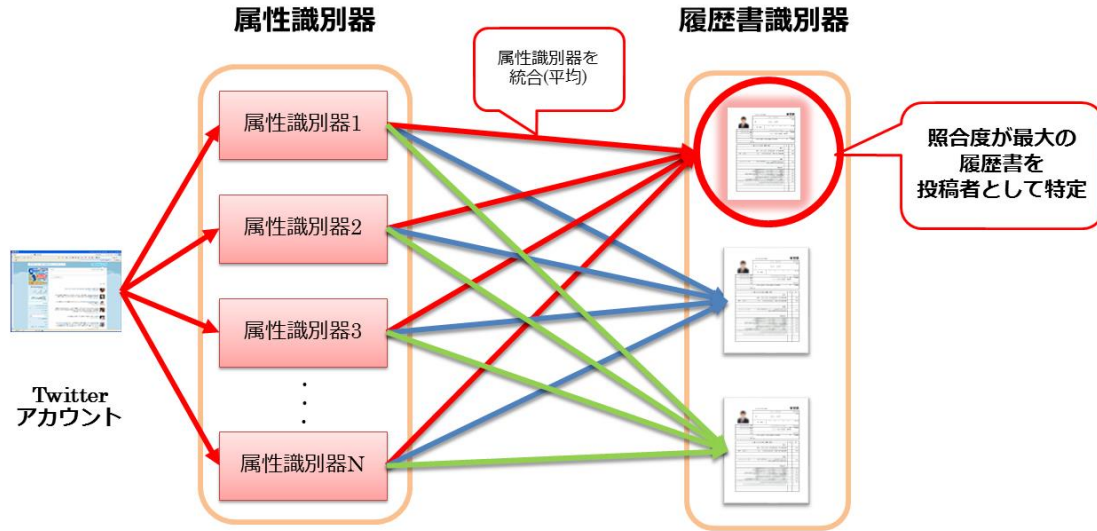


図 5.2 個人特定の手法

識別器毎の結果を補正するため、出力結果全てに対して次の式により正規化を行った。本人が未知の Twitter アカウント i に対して各識別器を j としたとき、正規化後の値 α_{ij} は

$$\alpha_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad \dots \text{式 5.1}$$

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij} \quad \dots \text{式 5.2}$$

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2}{N}} \quad \dots \text{式 5.3}$$

- i : Twitter アカウント
- j : 履歴書の属性
- N : 履歴書の数
- x_{ij} : Twitter アカウント i の識別器 j に対するスコア
- \bar{x}_j : 識別器 j におけるスコアの平均
- σ_j : 識別器 j におけるスコアの標準偏差値

5.2 個人特定の結果と考察

スコア統合後の照合結果を以下に記す。

表 5.1 照合結果

| | 全体(78 人) | 学内(27 人) | 学外(51 人) |
|-------------|-------------|------------|-------------|
| 本人と特定できた人数 | 29 (1 位) | 12 (1 位) | 27 (1 位) |
| 1 割に絞り込めた人数 | 65 (7 位以内) | 16 (2 位以内) | 40 (5 位以内) |
| 2 割に絞り込めた人数 | 75 (15 位以内) | 24 (5 位以内) | 47 (10 位以内) |

図 5.3, 5.4 は箱ヒゲ図であり、横軸が被験者番号、縦軸が統合スコアの分布となっている。表中の箱は分布において 25%~75%の人物の値の範囲を示しており、箱から伸びている棒線(ヒゲ)は分布の最小値から最大値を示している。また、表中の記号について、「●」=本人のスコアが一番高い場合、「▲」=本人のスコアが 1 割以内の順位に相当する場合、「■」=本人のスコアが 2 割以内の順位に相当する場合、「×」=それ以外を表している。

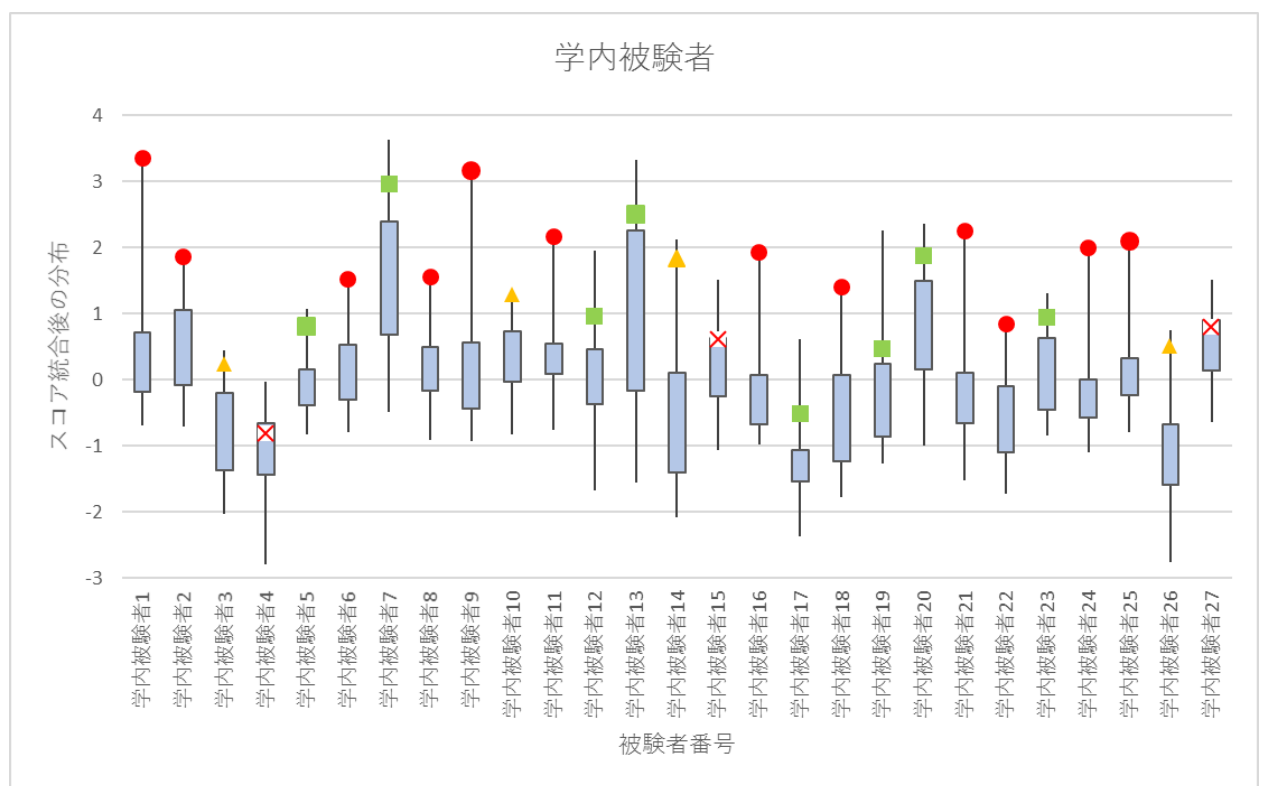


図 5.3 学内被験者の照合結果

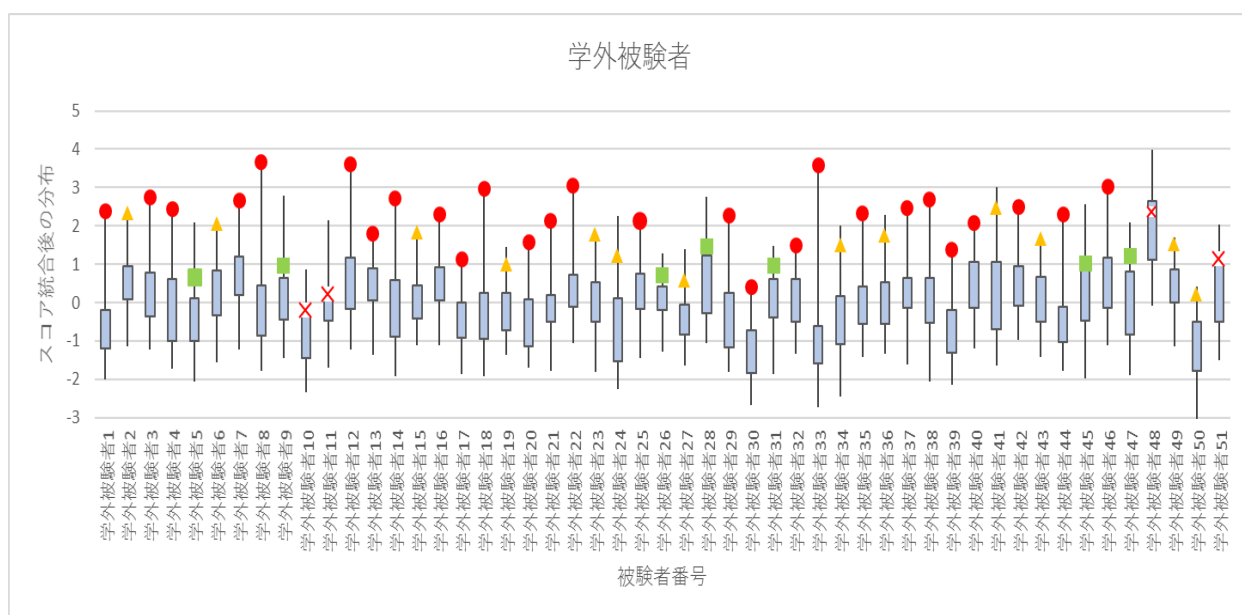


図 5.3 学外被験者の照合結果

5.3 推定されやすい人物

本人と特定できた被験者の傾向を分析することで、推定されやすい人物の特徴を分析する。

まず属性数に着目する。全体の中で本人を特定できた 29 人の属性数を集計したところ、一人当たりの属性の数は最低 6 項目、最大で 28 項目あり、平均は 11.41 項目となっている。3.1 で述べたように、全体における一人当たりの属性の数は最低で 6 項目、最大で 28 項目であり、平均は 11.14 項目となっている。以上より属性の数と推定のされやすさには相関はないと考えられる。

次に該当被験者の持つ属性の種類に着目する。初めに特定された被験者は共通の精度の高い属性を多く持っていると考え、29 人の履歴書に多く含まれていた属性を分析した。最も多かったのは「性別：男」で 21 人、次に「年齢：18～22」18 人、「得意科目：英語」10 人、「得意科目：数学」10 人と続いた。しかしこれらの属性について、4.2 節で述べた属性値識別器の精度計算における評価は高いものではなかった(表 5.2)。この理由として多くの被験者が持っている属性(=普遍的な属性)というものは識別精度が高くないものが多い。普遍的であるが故に Twitter に投稿されにくかったり、また属性値識別器の精度の計算が、該当する被験者が多いほど誤り率が高くなる方式であることなどが原因と思われる。

しかしここで 29 人の属性について計算をしたところ内 21 人が、共通ではないが誤り率 0 の属性の一つは持っていた(表 5.3)。つまり推定しやすい属性の一つでも持っていることがそのまま個人推定のされやすさにつながっているのではないかと考察した。誤り率 0 の属性とは 78 人の被験者において完全に本人のみを上位に特定できている属性である。そのため属性を保有している被験者が一人であるユニークな属性の場合が多い。

以上より、推定されやすい人物とは推定されやすい属性を持つ人物であり、さらに詳しい分析をするためには属性そのものについて調査する必要があると考える。該当する属性についての履歴書による申告などを控えることによって、Twitter アカウントから個人を推定されるのを防ぐことがある程度可能となるのではないかと。

表 5.2 履歴書に含まれている属性の個数と誤り率(一部)

| 属性 | 該当属性を履歴書に含む人数 (29 人内) | 誤り率 |
|-----------|--------------------------|----------|
| 男 | 21 | 12.3034 |
| 18-22(年齢) | 18 | 65.29993 |
| 英語 | 10 | 65.28833 |
| 数学 | 10 | 68.4 |
| 女 | 8 | 12.3034 |
| 23-26(年齢) | 6 | 55.03333 |
| 読書 | 5 | 49.83333 |
| 旅行 | 5 | 52.05242 |
| 英検 | 4 | 60.17143 |
| 漢検 | 4 | 33.58333 |
| 国語 | 4 | 57.44064 |

表 5.3 特定された被験者が持つ

誤り率 0 の属性の詳細(一部)

| | | 埼玉県A市 | 神奈川県B市 | 三重県C市 | D大学E学科 | 楽器F | スポーツG | スポーツH | ... | 合計 |
|-----------------|-------|-------|--------|-------|--------|-----|-------|-------|-----|----|
| 特定できた 被験者(※) | 被験者1 | ○ | | | | | | | | 1 |
| | 被験者6 | | ○ | | | | | | | 1 |
| | 被験者8 | | | | | | | | | 0 |
| | ... | | | | | | | | ... | |
| | 被験者28 | | | | ○ | | | | | 1 |
| | ... | | | | | | | | ... | |
| | 被験者31 | | | ○ | | | | | | 1 |
| | 被験者34 | | | | | ○ | ○ | | | 2 |
| | ... | | | | | | | | ... | |
| | 被験者65 | | | | | | | ○ | | 3 |
| (※)内の合計 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| 全体(78人)の合計 | | 1 | 1 | 2 | 1 | 1 | 1 | 1 | | |

第 6 章 隠れた属性の推定

6.1 目的

5 章の結論より被験者の属性について分析を続けたところ、本人が履歴書として申告していないが、本人の属性として高いスコアを出す属性が存在することが明らかになった。もしその属性が実際に本人の属性であるとしたら、この結果は申告時に被験者がその属性を自分のものとして自覚しなかったか、あるいは意図的に隠していたものではないかと考えられる。そのような潜在的な属性を見つけることができるか研究を行った。

6.2 実験方法

4 章で述べた属性識別器を用いることで、Twitter アカウントの投稿から投稿者が持っている潜在的な属性を推定することができると考え、実験を行った。

方法としては被験者の Twitter アカウントは投稿者が明らかであるため、投稿文のスコアが高い識別器の属性を投稿者に確認することで、推定できたかを判定した。この時、被験者が履歴書として事前に申告していない属性を確認することで、申告当時に本人が自覚していない、あるいは意図的に(履歴書として)申告しなかった属性を推定することが可能であるか否かを明らかにすることができる。

確認手法は、初めに推定する属性を趣味または得意科目に絞った。性別や住所についてはプライバシー性の高い情報であるため、履歴書以上のことについて詳しく追及するのを避ける形となった。その中でも、4.2 節で述べた属性値識別器の精度の計算において誤り率が小さいものから上位 25%である 51 属性を、信頼できる属性識別器として選択した。それらの属性の属性識別器から出力されたスコアについて、78 人分のスコアの中央値以上の値が出た属性を、推定された被験者の属性とした。

最後に、それらの属性についてアンケート方式で「非常に興味がある.」、「興味がある.」、「どちらでもない.」、「興味がない.」、「全く興味がない.」の 5 つの項目を用意し、被験者に回答してもらう(図 6.1)。選択された項目に応じてそれぞれ 5, 4, 3, 2, 1 のスコアを付け、本当に被験者の(潜在的な)属性であるか否かを調査する。このアンケートは被験者 78 人のうち、14 人から回答を得ることができた。

具体的な統計や分析結果については後述する。

SNSプライバシー保護の被験者実験（追加実験）

項目に対してどの程度の興味があるかを5段階で回答してください。

仮想通貨

必須

☐ 5. 非常に興味がある。
☐ 4. 興味がある。
☐ 3. どちらでもない
☐ 2. 興味がない。
☐ 1. 全く興味がない。

バレーボール

必須

☐ 5. 非常に興味がある。
☐ 4. 興味がある。
☐ 3. どちらでもない
☐ 2. 興味がない。
☐ 1. 全く興味がない。

フットサル

☐ 5. 非常に興味がある。

図 6.1 実際に用いたアンケートの例

6.3 推定の結果と考察

4 章で述べた属性識別器から出力されたスコアについて、78 人分のスコアの中央値以上の値が出た属性を、推定される被験者の属性とした。

表 6.1 を用いて説明する。ある属性 1 の識別器から算出されたスコアについて、被験者 78 人の中央値は 0.077878 であった。被験者 1、被験者 2、被験者 3 において、属性 1 の識別器スコアは全て中央値より高い。よってこの三人は属性 1 という属性を持っていると判断する。同様に属性 2 の識別器の結果より、被験者 1 と被験者 3 は属性 2 という属性を持っているが、被験者 2 は持っていないと判断する。同様に 51 属性、14 人についてツイッターの投稿文から推定される属性を調査し、アンケートの回答と比較した。

表 6.1 属性値識別器のスコアとその中央値(一部)

| | 属性1 | 属性2 | 属性3 | 属性4 | 属性5 |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 被験者1 | 0.935505 | 0.991285 | 0.764065 | 0.184813 | 0.978676 |
| 被験者2 | 0.127395 | 0.202059 | 0.578848 | 0.605516 | 0.055634 |
| 被験者3 | 0.245278 | 0.532414 | 0.02793 | 0.102823 | 0.563021 |
| | ・ | ・ | ・ | ・ | ・ |
| | ・ | ・ | ・ | ・ | ・ |
| | ・ | ・ | ・ | ・ | ・ |
| 中央値(78人) | 0.077878 | 0.295837 | 0.163126 | 0.275047 | 0.014716 |

アンケートから得られたスコアの平均を以下に示す(表 6.2). 左の列から順に, 51 属性全てのアンケートスコアの平均, 本人が申告している属性のアンケートスコアの平均, 本人が申告していないが属性値識別器より本人の属性と推定された(中央値以上のスコア)属性のアンケートスコアの平均, 本人の属性でないと推定された属性のアンケートスコアの平均となっている. 本人が申告しているスコアについては, 選択した 51 属性のみについてであるため, 該当するスコアがない被験者もいる.

回答を得られた被験者 14 人のうち, 被験者 2, 被験者 3, 被験者 5, 被験者 7, 被験者 12, 被験者 17, 被験者 19, 被験者 21, 被験者 22, 被験者 25 の 10 人については識別器から本人の属性と推定された属性のアンケートスコアが高いため, 被験者の属性として正しいものを推定できている.

表 6.2 アンケートスコアの平均

| | 51属性 | 本人の属性(申告有) | 本人の属性(申告無) | 本人の属性でない | 本人の属性の合計数 |
|-------|-------------|------------|-------------|-------------|-----------|
| 被験者1 | 2.470588235 | 5 | 2.434782609 | 2.2 | 26 |
| 被験者2 | 2.725490196 | 4 | 2.76 | 2.64 | 26 |
| 被験者3 | 2.921568627 | | 3.181818182 | 2.444444444 | 33 |
| 被験者5 | 1.941176471 | | 2.074074074 | 1.791666667 | 27 |
| 被験者7 | 2.568627451 | 5 | 2.619047619 | 2 | 43 |
| 被験者9 | 2.843137255 | | 2.846153846 | 2.846153846 | 26 |
| 被験者12 | 3.196078431 | | 3.34375 | 2.947368421 | 32 |
| 被験者17 | 3.078431373 | | 3.666666667 | 2.833333333 | 15 |
| 被験者19 | 2.431372549 | | 2.586206897 | 2.227272727 | 29 |
| 被験者20 | 2.843137255 | 5 | 2.625 | 2.961538462 | 25 |
| 被験者21 | 2.980392157 | 5 | 3.4 | 2.676470588 | 17 |
| 被験者22 | 2.450980392 | | 2.55 | 2.387096774 | 20 |
| 被験者23 | 2.784313725 | | 2.689655172 | 2.909090909 | 29 |
| 被験者25 | 2.31372549 | 5 | 2.454545455 | 2.107142857 | 23 |

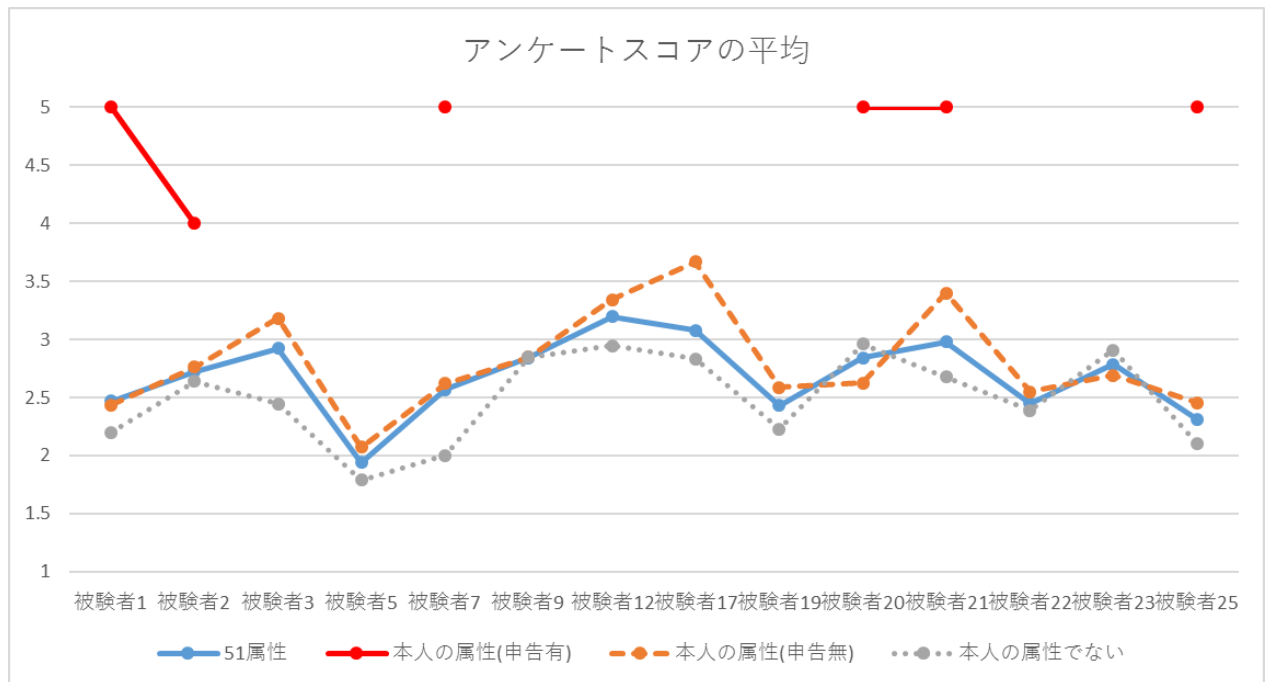


図 6.3 アンケートスコア

第7章 結論

7.1 まとめ

近年、SNS におけるプライバシーの侵害が問題となっている。プライバシーリスクを明確化するために、先行研究では同一人物による複数のアカウントを特定する手法が提案されている。本研究ではリスクをより明確化するために実世界の個人を特定する手法に焦点を当て、機械学習を用いて Twitter アカウントと就職希望者の履歴書を照合することで、匿名の発言の投稿者を見つける先行研究を取り上げた。この先行研究の信頼性を高めるとともに、SNS から投稿者が持つ属性をプロファイリングし、その精度を確認、分析することで推定されやすい人物の持つ性質を考察した。

プロファイリングについては、Twitter の投稿文から投稿者の属性を推定した。履歴書に記入された属性の中から、「性別」、「年代」、「帰省先住所」、「学歴」、「職種」、「得意科目」、「クラブ活動・サークル・趣味」、「資格」の 8 カテゴリーに対して属性推定を行った。教師データについてはデータセットを正例、負例共に最大 100 人と先行研究より増加させ、また被験者を 78 人に変更し、属性数を大幅に増やして精度を確認した。結果、最も識別精度の高い属性カテゴリは性別であり、学歴、帰省先住所と続いた。反対に最も識別精度の低い属性カテゴリは年代であり、資格、得意科目と続いた。ここで複数の被験者について、履歴書に記入されていなかった属性の推定値が高かったことから、投稿者の潜在的な属性の推定についての可能性を示した。

投稿者の隠れた属性の推定については、申告されていないが機械学習によって本人のものと推定された属性が実際に正しいのかを分析し、それにより潜在的な属性を見つけることができる可能性を示した。機械学習による属性推定とアンケートによる事後申告により、14 人中 10 人については事前に申告がない属性についても、本人のものと推定された属性のアンケートスコアが実際に高くなっていた。

個人特定については、78 人中 29 人を正しく特定し、78 人中 65 人を 1 割(すなわち上位 7 位)以内に絞り込むことができた。これは先行研究の結果(30 人中 6 人を正しく特定、30 人中 12 人を 1 割以内に特定)を上回る。また学内被験者では 27 人中 12 人を正しく特定し、27 人中 16 人を 1 割(すなわち上位 2 位)以内に、学外被験者では 51 人中 27 人を正しく特定し、51 人中 40 人を 1 割(すなわち上位 5 位)以内に絞り込むことができた。

7.2 今後の課題

推定されやすい人物や潜在的な属性についてはさらに詳しい分析が必要である。今後はこれらの分析から、反対にどのような投稿が推定されにくいのかを考察し、それを踏まえて匿名化を実施、情報の漏洩を防ぐ手法について考案していく。

謝辞

本研究の遂行ならびに本論文の作成にあたって、細部に至るまで終始ご指導とご鞭撻を賜った吉浦裕教授をはじめ、本研究において丁寧なご指導を頂いた内海彰教授、市野将嗣教授、学習院大学の久保山哲司教授に深く感謝いたします。

また、本研究の遂行ならびに本論文の作成に当たって様々なご指導ご助言をいただきました吉浦・市野研究室の皆様に深く感謝致します

参考文献

- [1] Narayanan, A., Shmatikov, V.: De-anonymizing social networks, In: Proc. 30th IEEE Security & Privacy, pp.173-187, Oakland (2009)
- [2] Eina Hashimoto, Masatsugu Ichino, Tetsuji Kuboyama, Isao Echizen, and Hiroshi Yoshiura, Breaking Anonymity of Social Network Accounts by Using Coordinated and Extensible Classifiers based on Machine Learning (2016)
- [3] Mao, H., Shuai, X., Kapadia, A.: Loose Tweets: An analysis of privacy leaks on Twitter. In: Proceedings of 10th ACM Workshop on Privacy in the Electronic Society, Denver (2011)
- [4] Kótyuk, G., Buttyan, L.: A Machine learning based approach for predicting undisclosed attributes in social networks. In: Proceedings of IEEE 4th International Work-shop on Security and Social Networking, pp.361–366, Budapest (2012)
- [5] A. Narayanan, et al., “On the Feasibility of Internet-Scale Author Identification,” in Proceedings of the 33rd IEEE Symposium on Security and Privacy, pp.300–314(2012)
- [6] Polakis, et al., “Using Social Networks to Harvest Email Addresses,” in Proceedings of the 9th ACM Workshop on Privacy in Electronic Society, pp.11–20(2010)
- [7] Goga, et al., “On Exploiting Innocuous User Activity for Correlating Accounts Across Social Network Sites,” ICSI Technical Reports -University of Berkeley, (2012).
- [8] A. Narayanan, et al., “On the Feasibility of Internet-Scale Author Identification,” in Proceedings of the 33rd IEEE Symposium on Security and Privacy, pp.300–314(2012)
- [9] Mishari Almishari et al, “Stylometric Linkability of Tweets,” in Workshop on Privacy in the Electronic Society, USA (2014)
- [10] ツイプロ - <http://twpro.jp/>, 2019/1/26 確認
- [11] DMLC for Scalable and Reliable Machine Learning -<http://dmlc.ml/>, 2019/1/26 確認
- [12] Scalable, Portable and Distributed Gradient Boosting (GBDT, GBRT or GBM)

Library, for Python, R, Java, Scala, C++ and more. -<https://github.com/dmlc/xgboost>,
2019/1/26 確認

付録

表 4.3 に示した属性識別器誤り率の全体を記す. なお被験者の個人情報が含まれるため一部を非公開とする.

| 属性値 | | 誤り率 |
|-------|------------|-------------|
| 性別 | 男 | 12.30339623 |
| | 女 | 12.30339623 |
| 年代 | 18-22 | 65.29993408 |
| | 23-26 | 55.03333333 |
| | 27-30 | 61.11780822 |
| | 31-35 | 42.73972603 |
| | 36-40 | 37.44 |
| | 41-45 | 23.50684932 |
| | 46-50 | 74.96103896 |
| 帰省先住所 | 愛知県**市 | 0 |
| | 愛媛県**市 | 0 |
| | ██████████ | 36.46753247 |
| | ██████████ | 5.064935065 |
| | ██████████ | 39.50649351 |
| | ██████████ | 10.12987013 |
| | ██████████ | 0 |
| | 熊本県**市市 | 0 |
| | ██████████ | 6.077922078 |
| | ██████████ | 7.184210526 |
| | 埼玉県**市市 | 0 |
| | 埼玉県**市市 | 2.025974026 |
| | ██████████ | 57.74025974 |
| | ██████████ | 0 |
| | ██████████ | 0 |
| | ██████████ | 0 |
| | ██████████ | 1.012987013 |
| | 秋田県**市 | 49.63636364 |
| | ██████████ | 0 |
| | ██████████ | 19.23287671 |
| | ██████████ | 20.25974026 |

| | | |
|--|--------|-------------|
| | | 7.090909091 |
| | | 0 |
| | | 64.83116883 |
| | | 0 |
| | | 0 |
| | 静岡県**市 | 5.064935065 |
| | 静岡県**市 | 31.4025974 |
| | | 1.012987013 |
| | | 0 |
| | 千葉県**市 | 12.15584416 |
| | | 0 |
| | | 2.025974026 |
| | | 4.051948052 |
| | | 1.012987013 |
| | | 47.61038961 |
| | | 2.025974026 |
| | | 69.8961039 |
| | | 15.19480519 |
| | 東京都**市 | 11.28947368 |
| | 東京都**市 | 10.26315789 |
| | 東京都**市 | 70.81578947 |
| | | 47.21052632 |
| | | 31.81578947 |
| | | 14.18181818 |
| | | 5.064935065 |
| | | 39 |
| | | 15.19480519 |
| | | 39 |
| | 栃木県**市 | 1.012987013 |
| | 栃木県**市 | 40.02631579 |
| | | 0 |
| | | 1.012987013 |
| | | 0 |
| | | 69.8961039 |
| | 北海道**市 | 0 |

| | | |
|----|--------------------|-------------|
| | | 6.077922078 |
| 学歴 | | 9.236842105 |
| | | 58.75324675 |
| | | 7.090909091 |
| | | 0 |
| | | 1.012987013 |
| | | 26.33766234 |
| | | 16.20779221 |
| | | 4.051948052 |
| | | 41.53246753 |
| | | 43.10526316 |
| | 電気通信大学情報理工学域 II 類 | 28.21764706 |
| | 電気通信大学情報理工学部総合情報学科 | 32.70967742 |
| | | 10.12987013 |
| | | 1.012987013 |
| | | 0 |
| | | 9.236842105 |
| | | 0 |
| | | 0 |
| | | 0 |
| | | 3.038961039 |
| 職種 | | 1.012987013 |
| | | 0 |
| | | 42.64 |
| | | 2.025974026 |
| | | 39 |
| | | 52.67532468 |
| | サービス業 | 60.32 |
| | | 46.5974026 |
| | | 56.72727273 |
| | 一般事務 | 50.53521127 |
| | 営業 | 26.33766234 |
| | | 1.012987013 |
| | | 23.2987013 |
| | | 37.97368421 |

| | | |
|----|------|-------------|
| | | 37.48051948 |
| | | 42.54545455 |
| | | 67.87012987 |
| | | 67.87012987 |
| | | 46.5974026 |
| | | 59.52631579 |
| | | 24.31168831 |
| | | 30.38961039 |
| | | 24.31168831 |
| | | 1.012987013 |
| | | 33.86842105 |
| | | 60.32 |
| 資格 | | 17.22077922 |
| | | 44.13157895 |
| | | 25.32467532 |
| | | 2.025974026 |
| | | 3.038961039 |
| | 英検 | 60.17142857 |
| | | 8.103896104 |
| | 漢検 | 33.58333333 |
| | 教員免許 | 60.55263158 |
| | | 8.103896104 |
| | | 31.4025974 |
| | | 66.85714286 |
| | | 39.52 |
| | | 3.038961039 |
| | | 36.46753247 |
| | | 41.53246753 |
| | | 40.51948052 |
| | | 54.7012987 |
| | | 67.87012987 |
| | | 63.81818182 |
| | | 62.80519481 |
| | | 42.54545455 |
| | 簿記二級 | 29.37662338 |

| | | |
|-------|------------|-------------|
| | | 38.49350649 |
| 得意科目 | | 22.28571429 |
| | Web セキュリティ | 5.064935065 |
| | | 64.65789474 |
| | | 0 |
| | | 22.57894737 |
| | | 37.6109589 |
| | | 0 |
| | | 6.077922078 |
| | | 3.038961039 |
| | | 9.116883117 |
| | | 46.90540541 |
| | | 9.116883117 |
| | 国語 | 57.44064386 |
| | | 56.72727273 |
| | 数学 | 68.4 |
| | 世界史 | 27.04 |
| | | 66.08333333 |
| | | 56.44736842 |
| | 通信システム | 61.79220779 |
| | | 69.78947368 |
| | | 46.18421053 |
| | | 5.131578947 |
| | | 41.05263158 |
| | | 72.93506494 |
| | | 51.66233766 |
| | | 36.46753247 |
| | | 1.012987013 |
| | | 13.16883117 |
| | | 65.28833173 |
| 特技・趣味 | | 1.012987013 |
| | | 74.96103896 |
| | | 39.50649351 |
| | アーチェリー部 | 0 |
| | | 5.064935065 |

| | | |
|--|---------|-------------|
| | アウトドア | 2.025974026 |
| | ■■■■■ | 61.79220779 |
| | ■■■■■ | 27.35064935 |
| | ■■■■■ | 44.29285714 |
| | アマチュア無線 | 4.051948052 |
| | ■■■■■ | 1.012987013 |
| | ■■■■■ | 4.051948052 |
| | ■■■■■ | 21.55263158 |
| | ■■■■■ | 42.54545455 |
| | ■■■■■ | 7.090909091 |
| | ■■■■■ | 29.12 |
| | ■■■■■ | 47.21052632 |
| | ■■■■■ | 21.08108108 |
| | ■■■■■ | 5.064935065 |
| | ■■■■■ | 7.090909091 |
| | ■■■■■ | 7.090909091 |
| | ■■■■■ | 45.15789474 |
| | ■■■■■ | 43.10526316 |
| | ■■■■■ | 20.8 |
| | ■■■■■ | 9.116883117 |
| | ギター | 14.56 |
| | ■■■■■ | 29.76315789 |
| | ■■■■■ | 0 |
| | ■■■■■ | 23.2987013 |
| | ■■■■■ | 5.064935065 |
| | ■■■■■ | 48.23684211 |
| | ■■■■■ | 5.064935065 |
| | ■■■■■ | 7.184210526 |
| | ■■■■■ | 4.051948052 |
| | ■■■■■ | 48.62337662 |
| | ■■■■■ | 13.16883117 |
| | ■■■■■ | 0 |
| | ■■■■■ | 10.12987013 |
| | ■■■■■ | 27.35064935 |
| | ショッピング | 55.42105263 |

| | | |
|--|----------|-------------|
| | | 55.71428571 |
| | | 5.27027027 |
| | | 10.12987013 |
| | | 0 |
| | | 7.184210526 |
| | | 26.68421053 |
| | テニス | 7.28 |
| | | 28.08 |
| | | 1.012987013 |
| | | 0 |
| | | 7.090909091 |
| | | 5.064935065 |
| | | 44.57142857 |
| | | 4.051948052 |
| | バスケットボール | 35.36 |
| | バレーボール | 4.051948052 |
| | | 3.038961039 |
| | | 1.012987013 |
| | | 12.15584416 |
| | | 32.14864865 |
| | ピアノ | 21.55263158 |
| | | 49.63636364 |
| | | 0 |
| | | 28.36363636 |
| | | 0 |
| | | 6.077922078 |
| | | 55.54545455 |
| | | 2.025974026 |
| | | 3.078947368 |
| | | 1.012987013 |
| | | 1.012987013 |
| | | 19.5 |
| | | 3.038961039 |
| | | 3.038961039 |
| | | 74.92105263 |

| | | |
|--|-------|-------------|
| | | 25.32467532 |
| | | 0 |
| | | 4.051948052 |
| | | 5.064935065 |
| | | 5.064935065 |
| | | 54.39473684 |
| | | 9.236842105 |
| | | 0 |
| | | 49.54054054 |
| | | 29.12 |
| | | 62.80519481 |
| | | 42.54545455 |
| | ラグビー部 | 1.012987013 |
| | | 1.012987013 |
| | | 73.94805195 |
| | | 30.38961039 |
| | ワイン | 4.051948052 |
| | | 0 |
| | | 48.62337662 |
| | | 74.96103896 |
| | | 71.92207792 |
| | | 13.34210526 |
| | | 13.52 |
| | | 6.157894737 |
| | | 1.012987013 |
| | | 1.012987013 |
| | | 24.96 |
| | | 36.46753247 |
| | | 44.13157895 |
| | 音楽鑑賞 | 70.62162162 |
| | | 4.051948052 |
| | | 51.66233766 |
| | | 44.13157895 |
| | | 35.45454545 |
| | | 18.72 |

| | | |
|--|-----|-------------|
| | | 3.038961039 |
| | | 3.038961039 |
| | | 2.025974026 |
| | | 0 |
| | 筋トレ | 45.32432432 |
| | | 58.75324675 |
| | | 17.22077922 |
| | | 2.025974026 |
| | | 16.20779221 |
| | | 40.51948052 |
| | | 26.33766234 |
| | | 31.4025974 |
| | | 33.42857143 |
| | | 54.81081081 |
| | | 1.012987013 |
| | | 0 |
| | | 30.98630137 |
| | | 74.88 |
| | | 21.27272727 |
| | | 26.33766234 |
| | 書道 | 11.14285714 |
| | | 9.236842105 |
| | | 15.6 |
| | | 6.077922078 |
| | | 50.59459459 |
| | | 12.15584416 |
| | | 27.35064935 |
| | | 23.2987013 |
| | | 13.16883117 |
| | | 0 |
| | | 15.19480519 |
| | | 4.051948052 |
| | | 26.33766234 |
| | | 0 |
| | | 1.012987013 |

| | | |
|--|-----|-------------|
| | | 0 |
| | | 3.038961039 |
| | | 19.24675325 |
| | | 0 |
| | | 6.077922078 |
| | | 6.077922078 |
| | 釣り | 7.184210526 |
| | | 57.74025974 |
| | 登山 | 2.025974026 |
| | | 31.4025974 |
| | | 4.051948052 |
| | | 49.83333333 |
| | | 49.26315789 |
| | | 5.064935065 |
| | | 2.025974026 |
| | | 5.064935065 |
| | | 0 |
| | | 27.35064935 |
| | | 3.038961039 |
| | | 13.34210526 |
| | | 37.39726027 |
| | 無線部 | 2.025974026 |
| | 野球 | 5.27027027 |
| | | 49.63636364 |
| | 旅行 | 52.05241935 |
| | 料理 | 63.23571429 |
| | | 49.26315789 |
| | | 29.76315789 |
| | | 1.012987013 |